Rolling Out GENESIS/SciFlo in the ESIP Federation's **Earth Information Exchange** (EIE)

Tom Yunck, Brian Wilson, Eric Fetzer, Michael Garay, Gerald Manipon, Benyang Tang, and Amy Braverman
Jet Propulsion Laboratory
tpy@jpl.nasa.gov

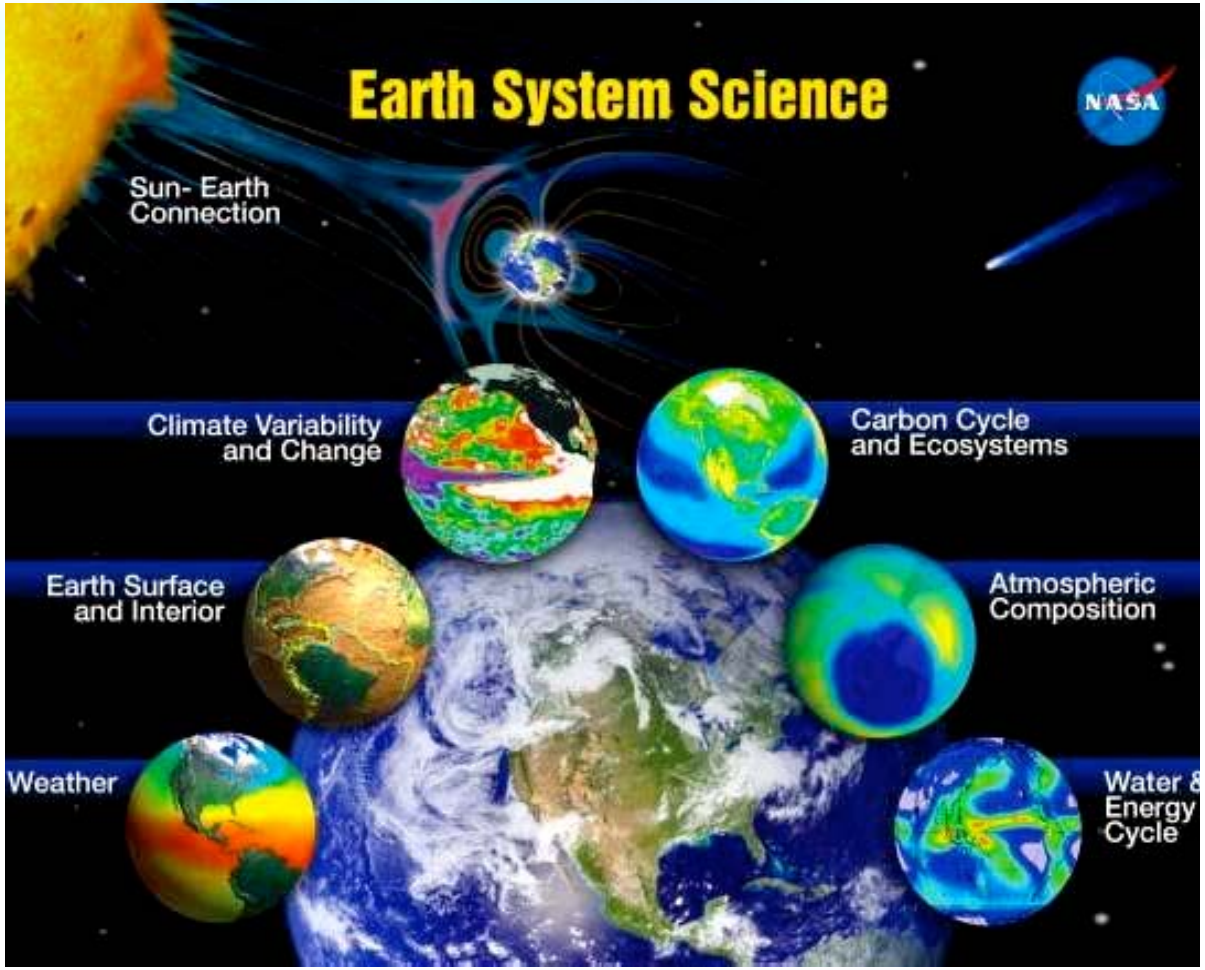
Definitions

GENESIS: **GEN**eral **E**arth **S**cience **I**nvestigation **S**uite — A REASoN project now in its 3rd year.

SciFlo: The workflow execution engine that makes it all work

Earth Information Exchange: The ESIP Federation's new Earth Science Information portal

The Vision of Earth System Science



- **Characterize**
  Earth's varied behavior
- **Understand**
  the Earth as an integrated system
- **Predict**
  Earth's response to complex forcings

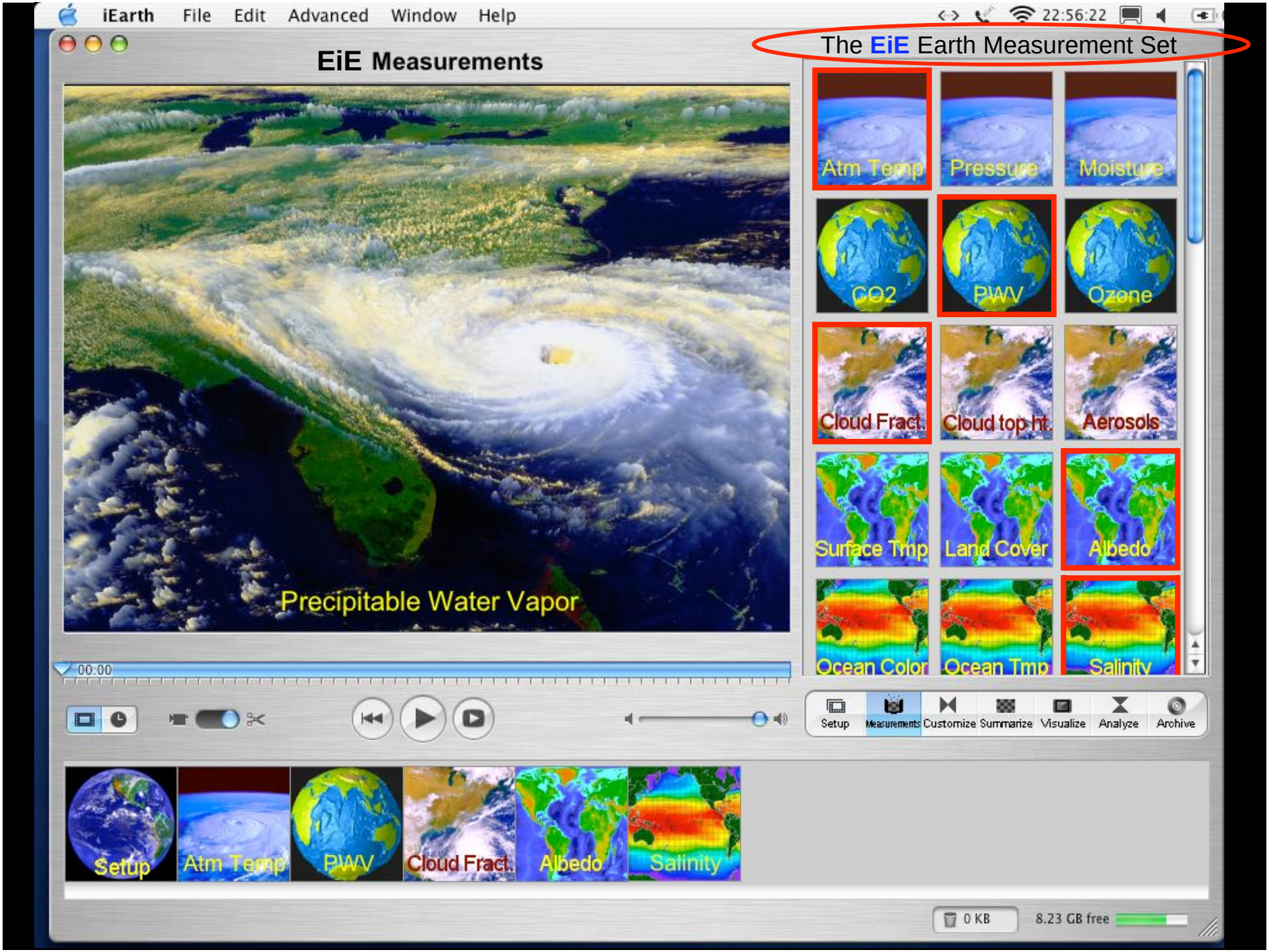
Earth Science IT Challenges

- Coping with diverse Earth science data sets:
  - Locating the right products (Data Discovery)
  - Selecting: browse, query, subset, customize…
  - Retrieving large data volumes swiftly
  - Fusing diverse, incommensurate products
  - Visualizing massive multidimensional data
  - Discovering knowledge: Summarize/Analyze/Mine/…
- Sample research scenario Today: Multi-year effort for a modest, cross-instrument study

A Sketch of the GENESIS Vision


Finder File Edit View Go Window Help
83:17:14
(Charged)
Tue 11:25 PM
iEarth


iEarth
File
Edit
Advanced
Window
Help
22:56:22
New Project... ⌘N
Open Project... ⌘O
Open Recent
Save Project ⌘S
Save Frame As... ⌘F
Import... ⇧⌘I
Export... ⇧⌘E
Show Info ⌘I
Empty Trash...
Layout
Please Begin
00:00
Setup
Measurements
Customize
Summarize
Visualize
Analyze
Archive
Drag clips here to build your project.
0 KB
8.23 GB free


iEarth File Edit Advanced Window Help
22:56:22
EiE Measurements
The EiE Earth Measurement Set
Precipitable Water Vapor
Atm Temp
Pressure
Moisture
CO2
PWV
Ozone
Cloud Fract.
Cloud top ht.
Aerosols
Surface Tmp
Land Cover
Albedo
Ocean Color
Ocean Tmp
Salinity
00:00
Setup
Measurements
Customize
Summarize
Visualize
Analyze
Archive
Setup
Atm Temp
PWV
Cloud Fract.
Albedo
Salinity
0 KB
8.23 GB free


iEarth
File
Edit
Advanced
Window
Help
22:56:22
EiE Analysis
Operators
00:00
Setup
Measurements
Customize
Summarize
Visualize
Analyze
Archive
Setup
Atm Temp
PWV
Cloud Fract
Albedo
Salinity
Customize
Summarize
Visualize
Analyze
0 KB
8.23 GB free

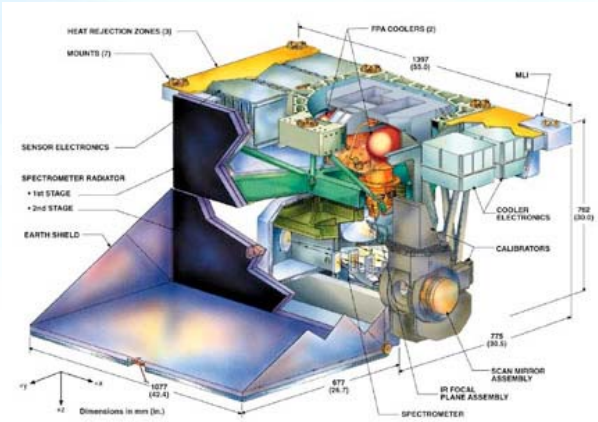
Five Core Ideas

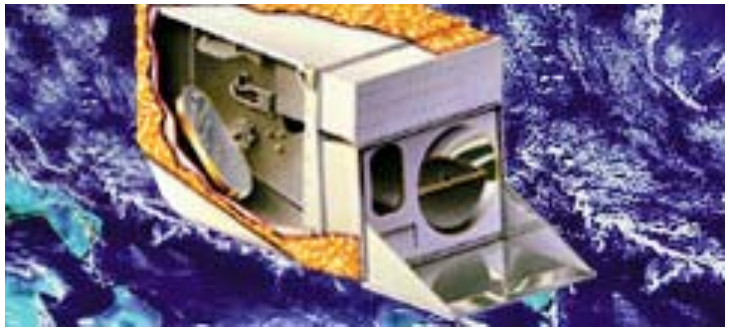
- Loosely-coupled distributed computing using SOAP services
- Exposing scientific analysis operators as SOAP web services
- Specifying a processing stream as an XML document
- Dataflow engine for a parallel execution and load balancing
- Visual point-and-click programming

General Example

Integrated space/time query and access for AIRS, MODIS, MISR and GPS-RO data

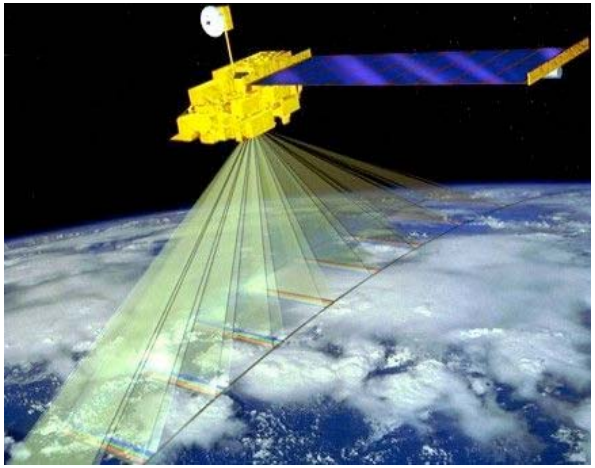
AIRS

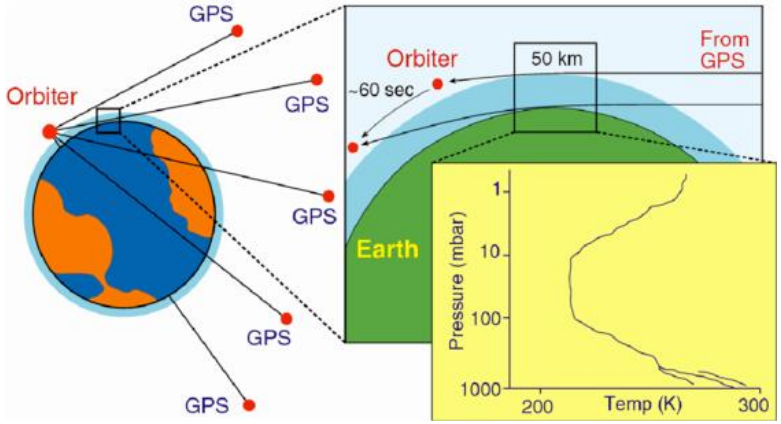




MODIS

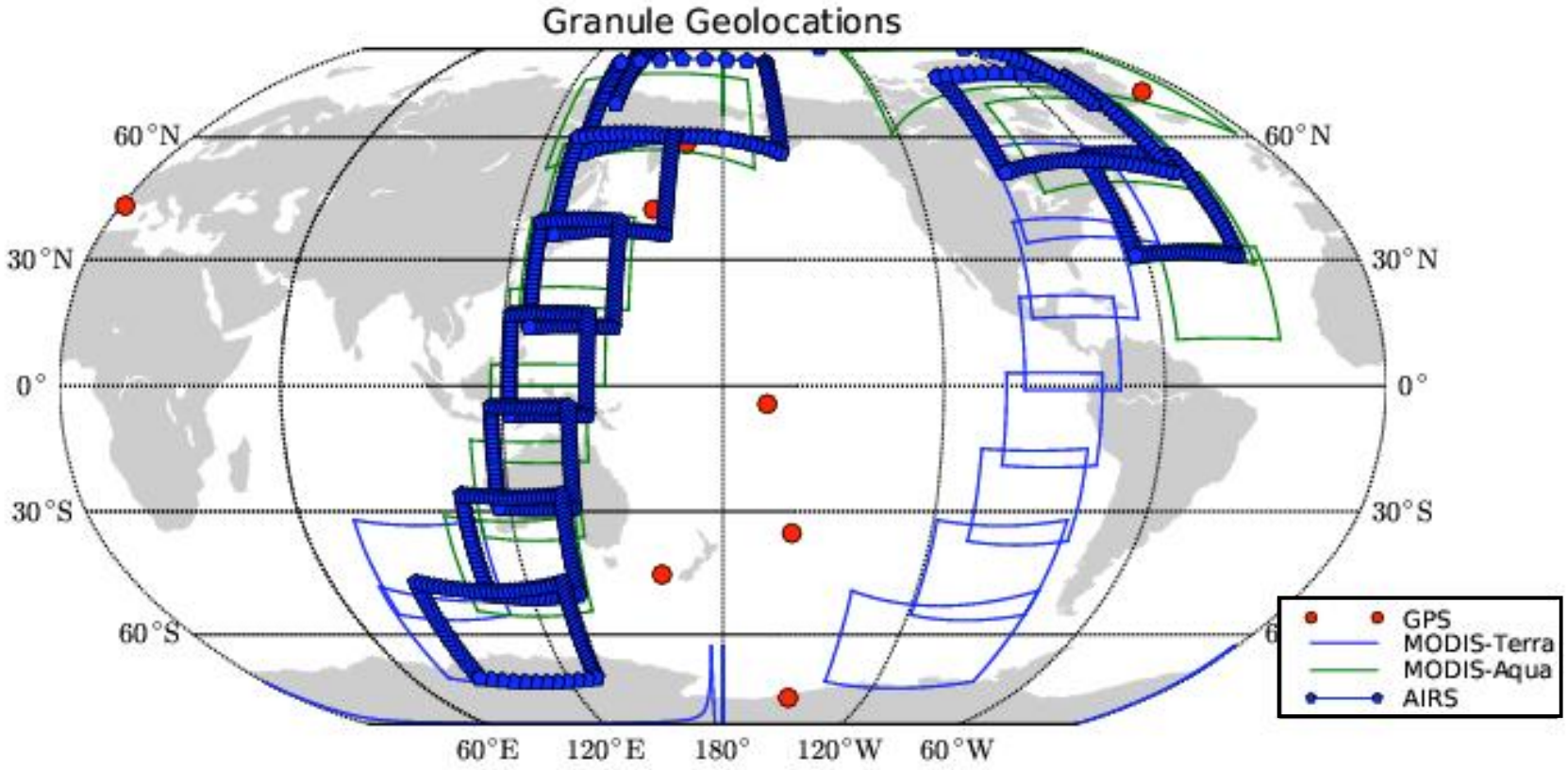
MISR





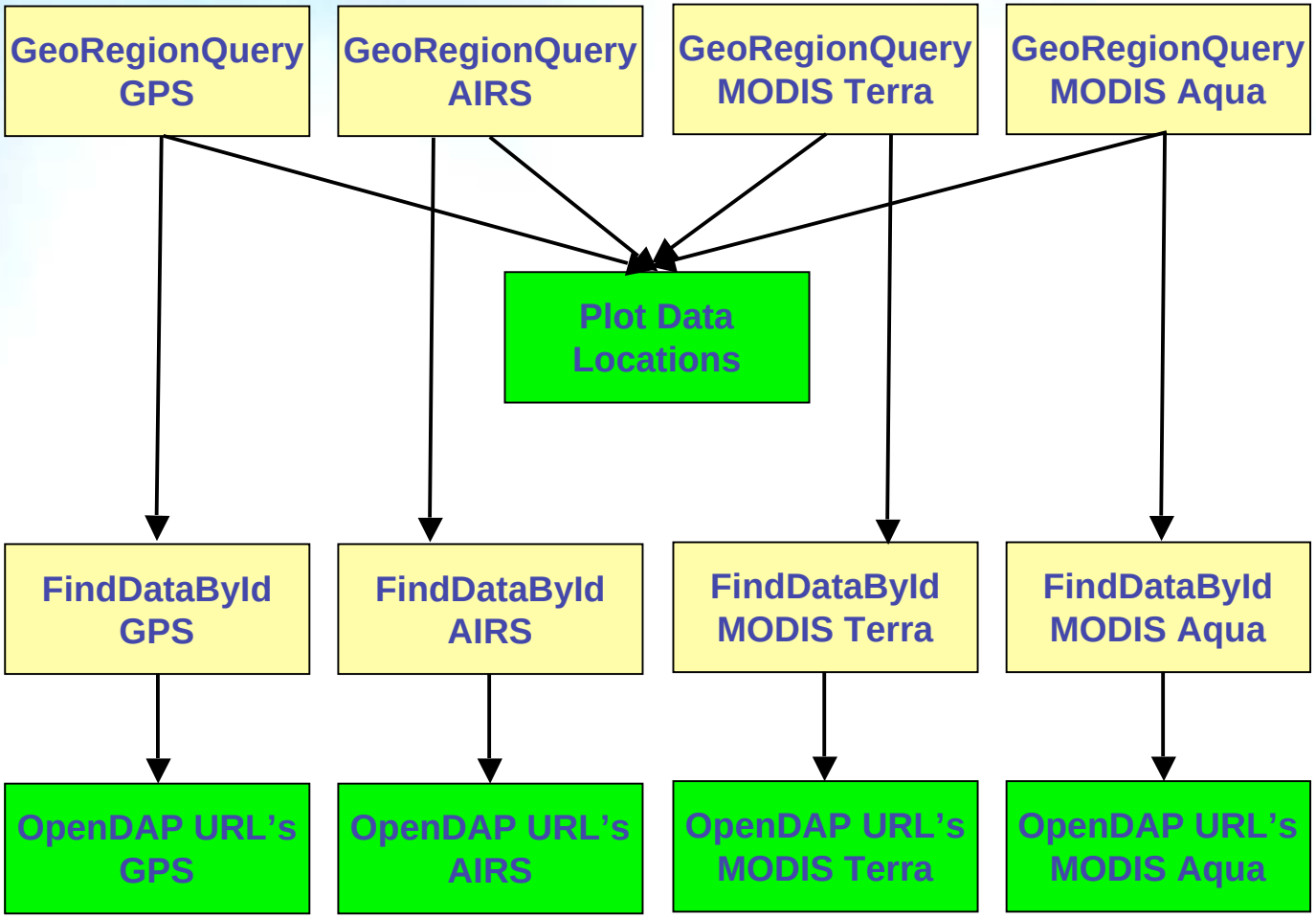
GPS Radio Occultation

Geolocation Plot for GPS, AIRS, MODIS



Try this SciFlo at http://sciflo.jpl.nasa.gov

GeoRegionQuery for GPS, AIRS, MODIS



The Current GENESIS/SciFlo Interface

you are at: home » Registered User's Area » Demo » GPS-AIRS Matchup v0.3b (GPS version 2.0)

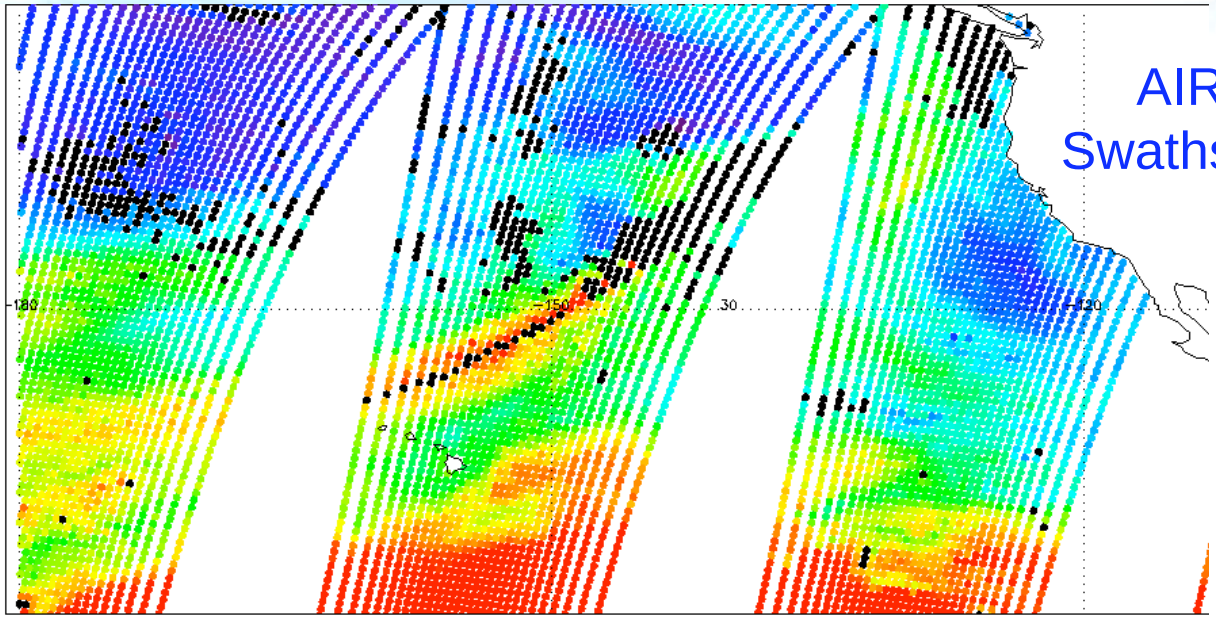
SESSION id: **1102910859060900171 1**

Starting Date/Time	2003 / 01 / 04 00 : 00
Ending Date/Time	2003 / 01 / 04 23 : 59
North Bounding Latitude	90.0
South Bounding Latitude	-90.0
West Bounding Longitude	-180.0
East Bounding Longitude	180.0
Time Tolerance (seconds)	7200
Location Tolerance (km)	200
Priority	1
Retrieval Type Max (between 0 and 100)	10
Land Fraction Min	0
Land Fraction Max	.1
Output whole swath?	☐

OK

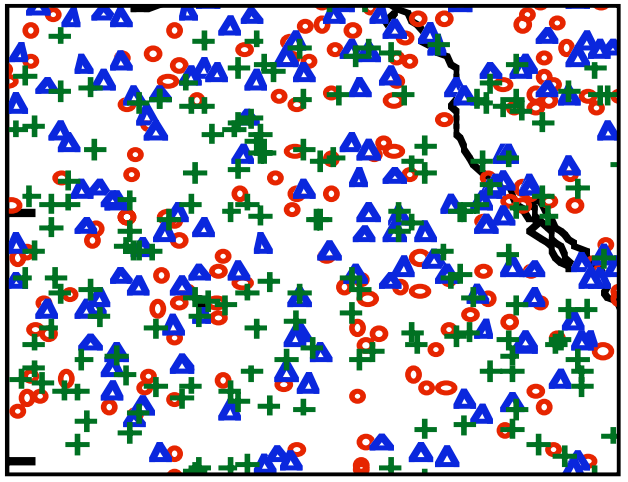
- Interface: HTML web form auto-generated from XML dataflow doc.
- Input: User enters start/end time and other co-registration criteria.
- Flow Execution: Calls 4 SOAP data query services and 15 operators on 4 computers.

AIRS/GPS Co-registration: Point-to-Swath

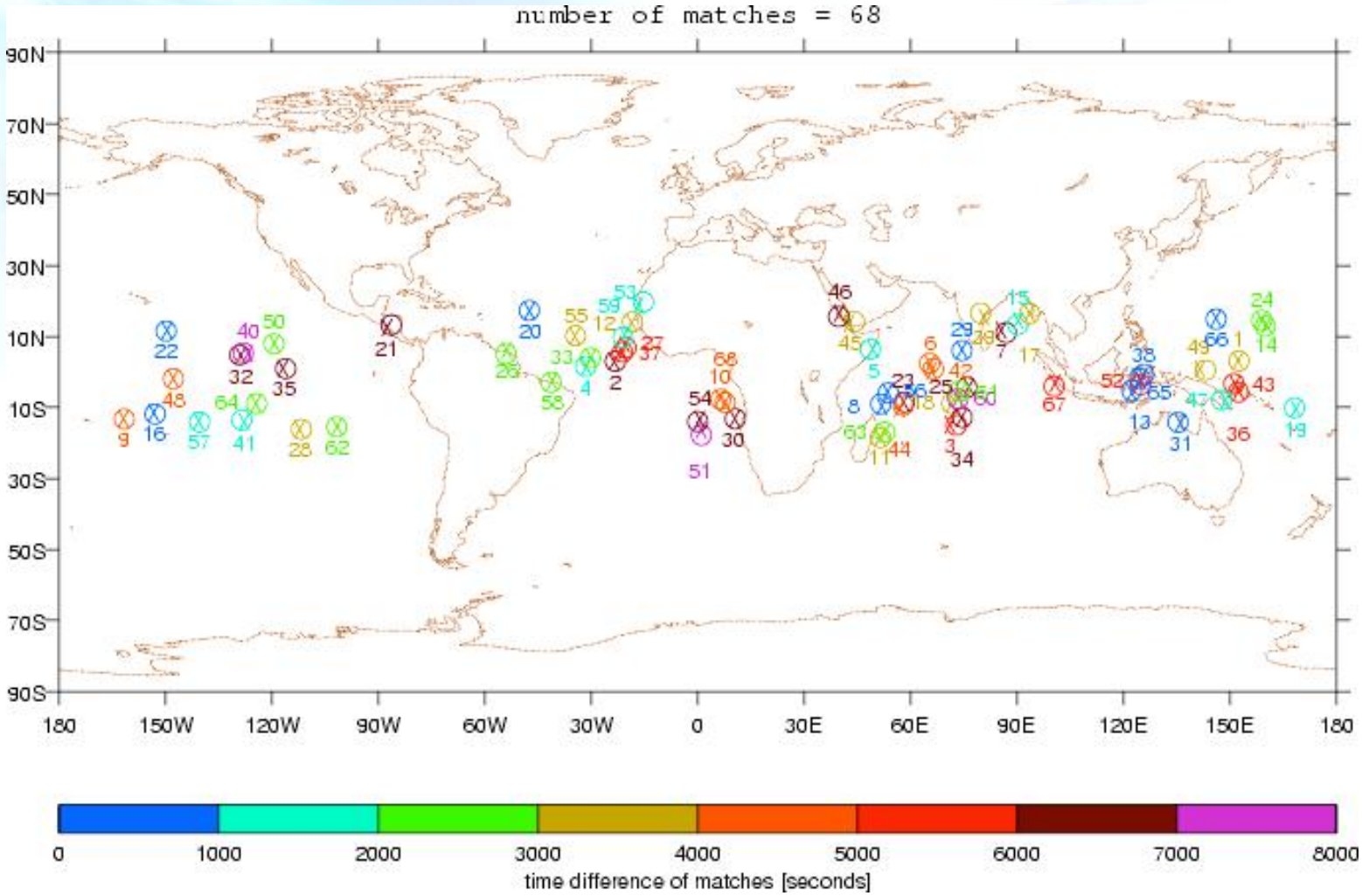


AIRS Level-2
Swaths over Pacific

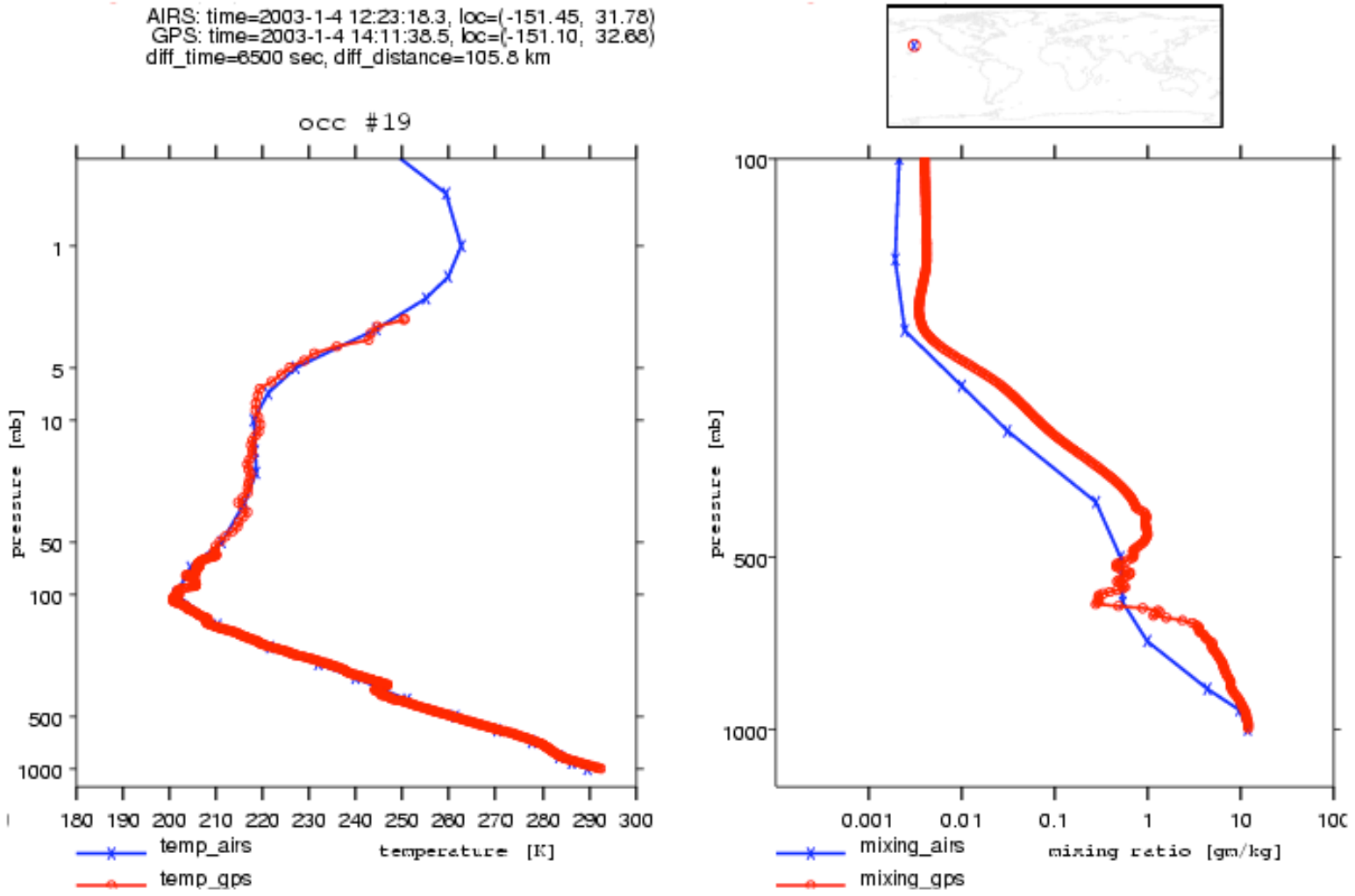
GPS Level-2
Profile Locations



AIRS/GPS Matchups

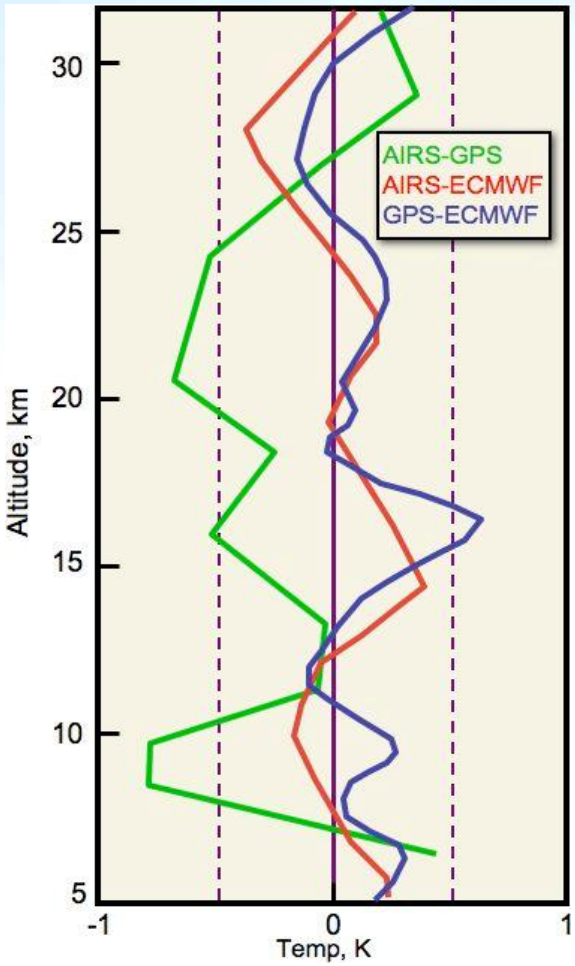


AIRS/GPS Temperature & Moisture Comparisons

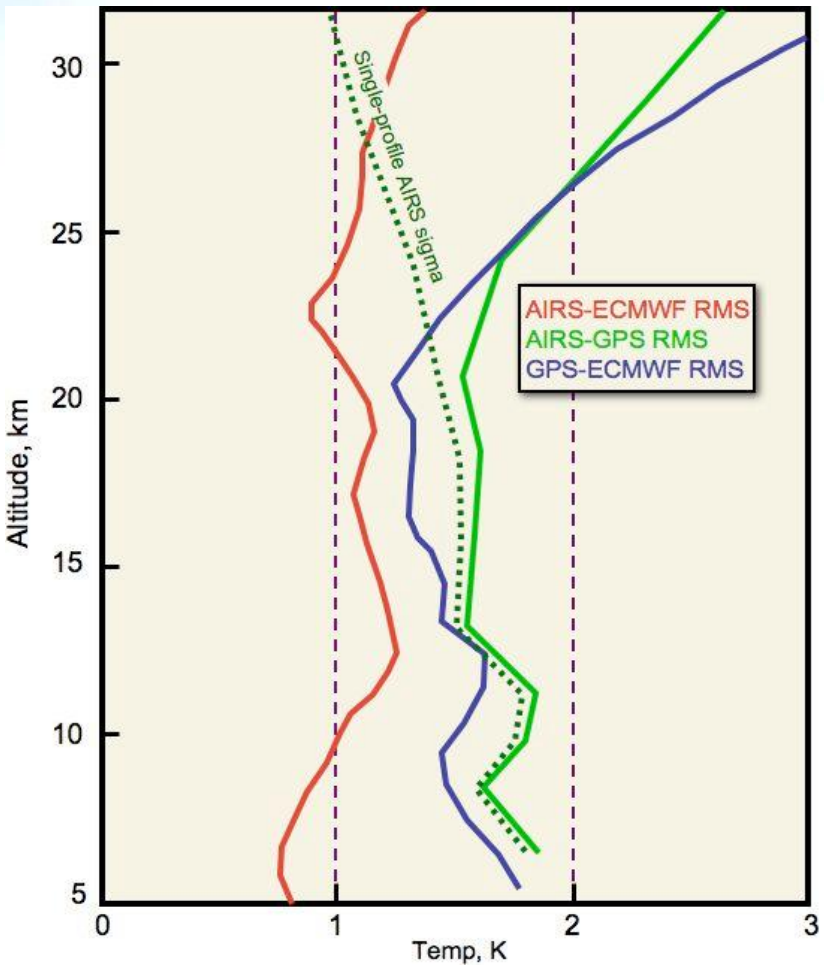


Sample Comparison Statistics

AIRS / GPS-RO / ECMWF Temperature Comparisons



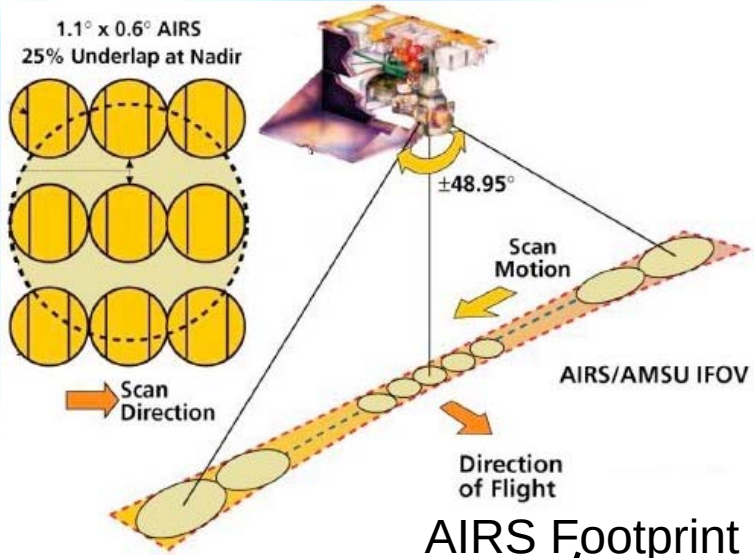
Mean Difference



RMS Difference

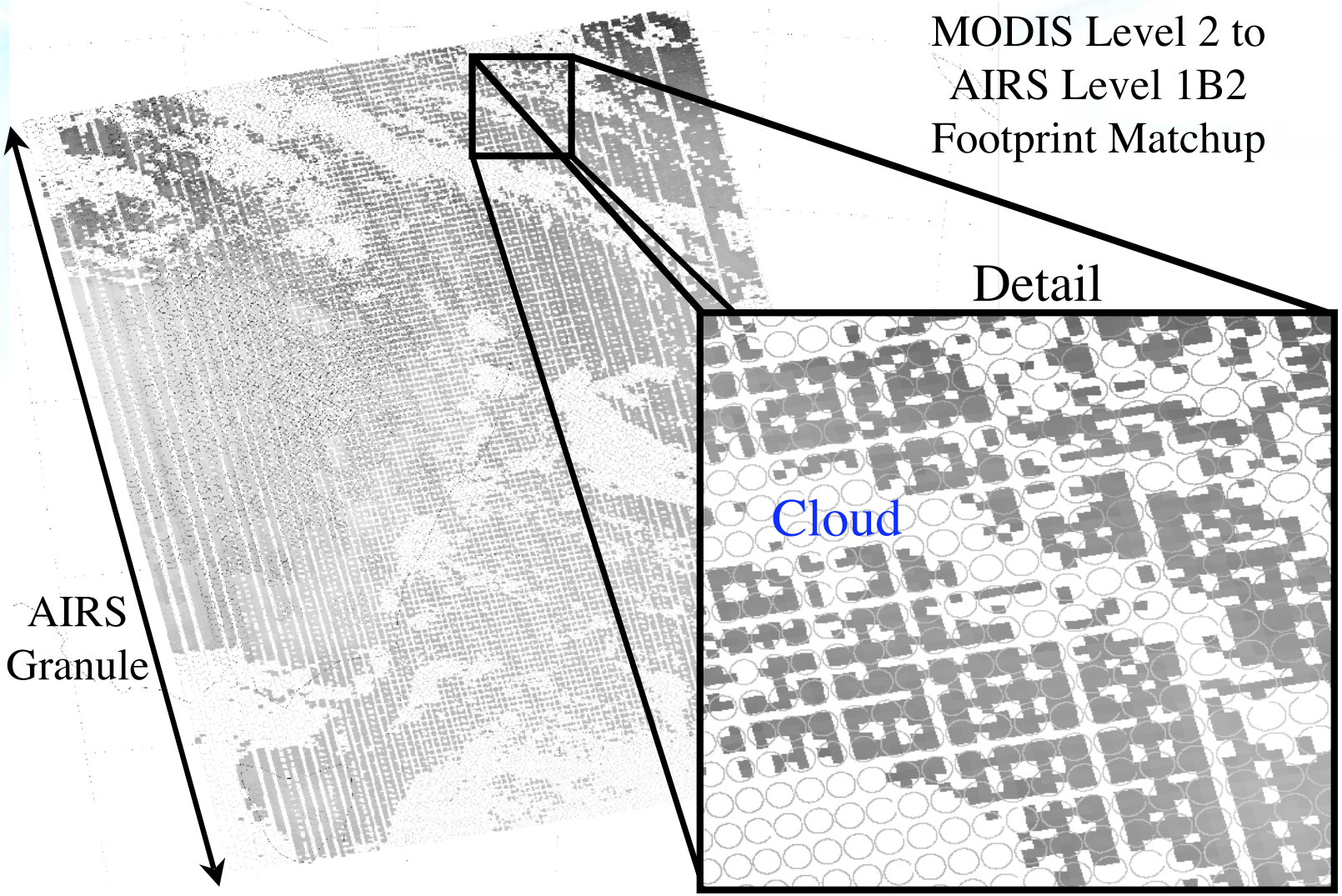
Yunck et al., 06/27/06

Example 3: AIRS/MODIS Matchups



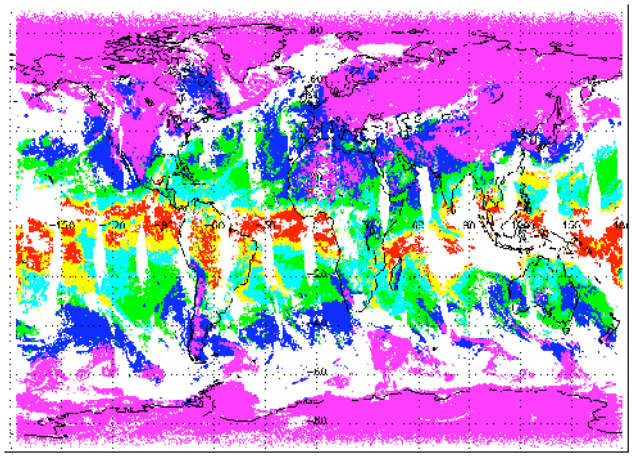
The AIRS field of view lacks overlap and increases with scan angle, requiring careful matching to MODIS footprints for consistent comparisons between retrievals from the two instruments.

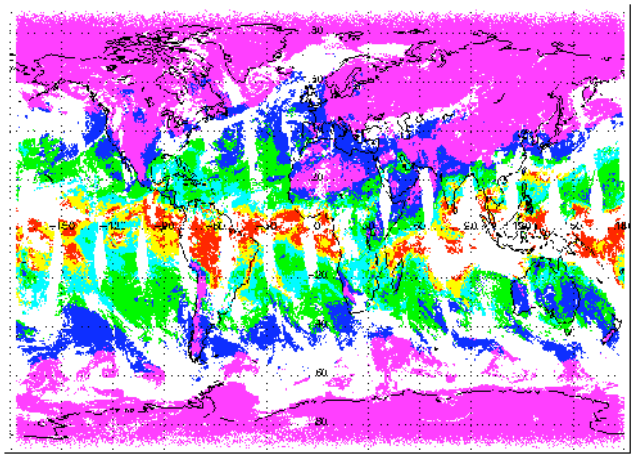
AIRS/MODIS Matchups (cont.)



AIRS/MODIS Moisture (TPW) Comparisons

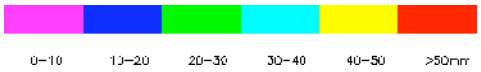
AIRS
Day/Night
Combined



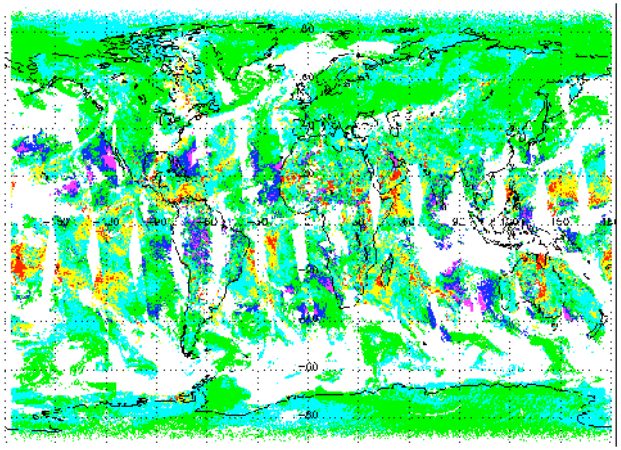


MODIS
Day/Night
Combined

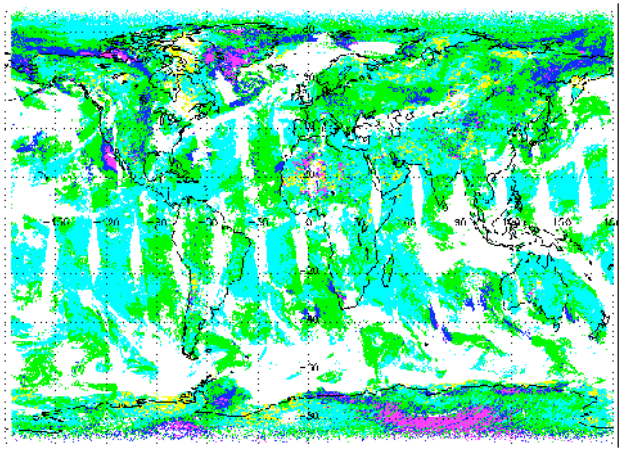
Total Precipitable Water (mm)



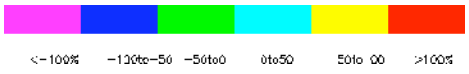
Bias
Day/Night
Combined



AIRS-MODIS Bias (mm)



AIRS-MODIS Percent Difference

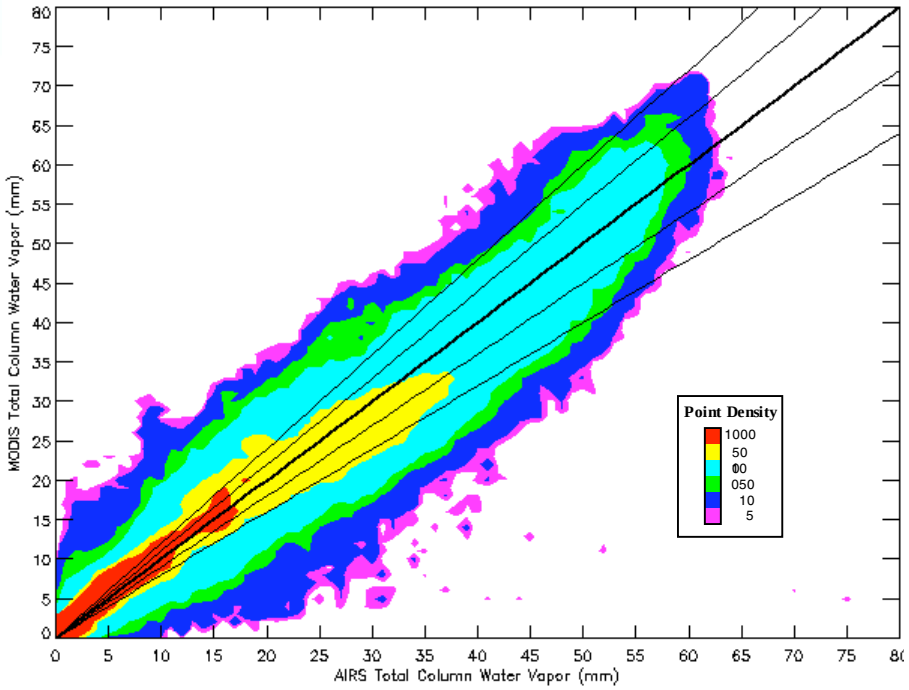


% Difference
Day/ Night
Combined

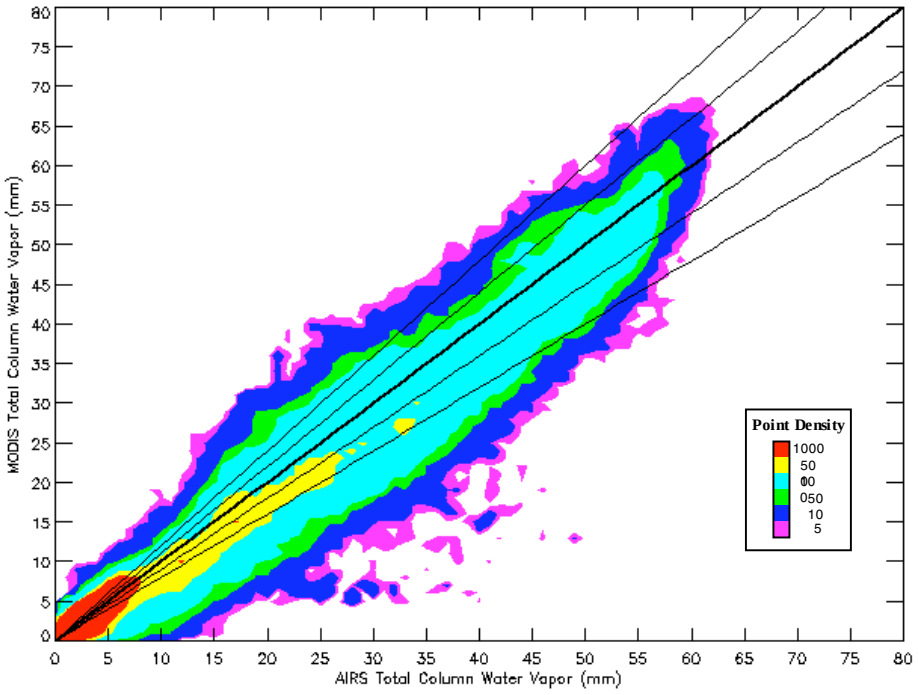
Yunck et al., 06/27/06

AIRS/MODIS Moisture Comparisons

MODIS vs. AIRS Total Column Water Vapor Density (Day)



MODIS vs. AIRS Total Column Water Vapor Density (Night)



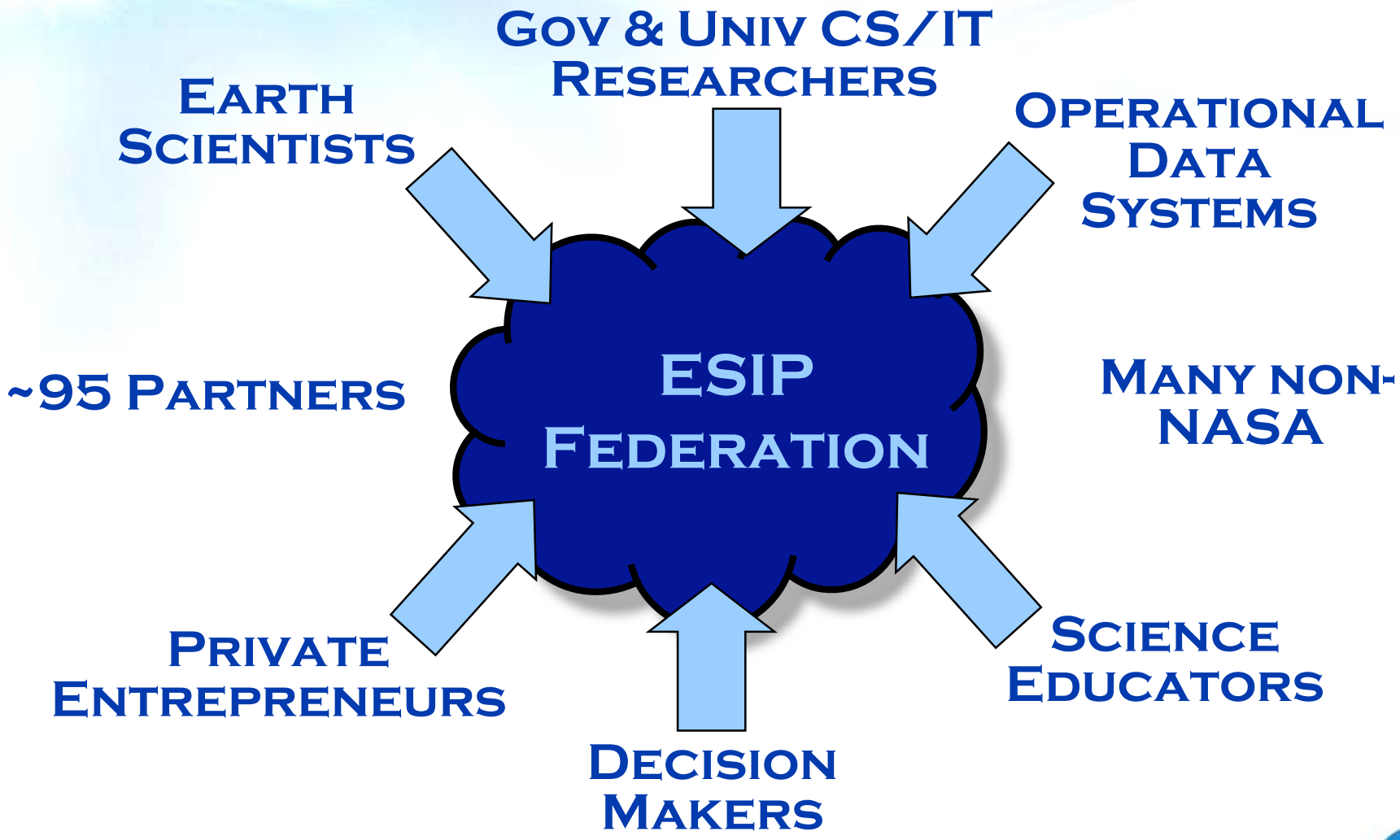
Yunck et al., 06/27/06

Flexible Distributed Computing with GENESIS/SciFlo

- Move data to the operators – Move entire AIRS granules from the Goddard DAAC to a local SciFlo server at JPL. Compare AIRS to GPS locally.
- Move operators to the data – Move the comparison operators and the flow execution into a SciFlo server that has local access to the large AIRS granule archive.
- Custom parameter subsetting – Subset the AIRS granules at the DAAC using a custom subsetting operator to generate tailored granules. Transfer the smaller files to a local SciFlo server.
- On-demand OpenDAP Slicing – Leave the AIRS granules in the Goddard Data Pool and perform the comparison at JPL. Using OpenDAP, only grab the handful of variables (temp, water vapor, etc.) from each granule that are needed for the comparison.


Earth Information on the Global Grid
The
ESiP
FEDERATION
Science • Applications • Decision Support

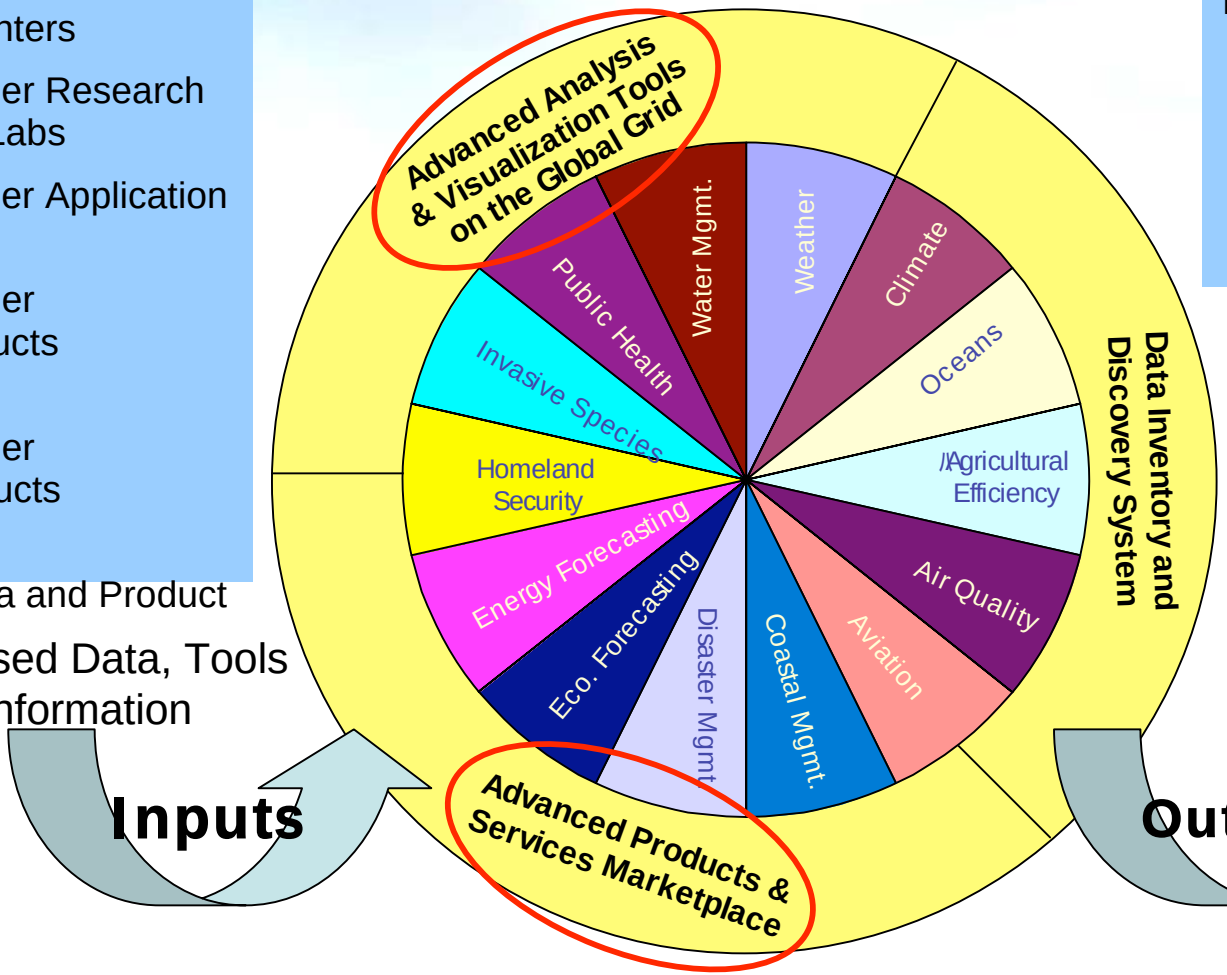
Rolling Out *GENESIS* in the Earth Information Exchange



The Earth Information Exchange *Portal*

- NASA, NOAA & USGS Member Data Centers
- Federation Member Research Universities and Labs
- Federation Member Application Developers
- Federation Member Technology Products Developers
- Federation Member Educational Products Developers
- Non-Member Data and Product Suppliers

Science-based Data, Tools and Information



Inputs

Outputs

- **Decision Makers**
- **GIS Users**
- **Teachers**
- **Researchers**
- **General Public**

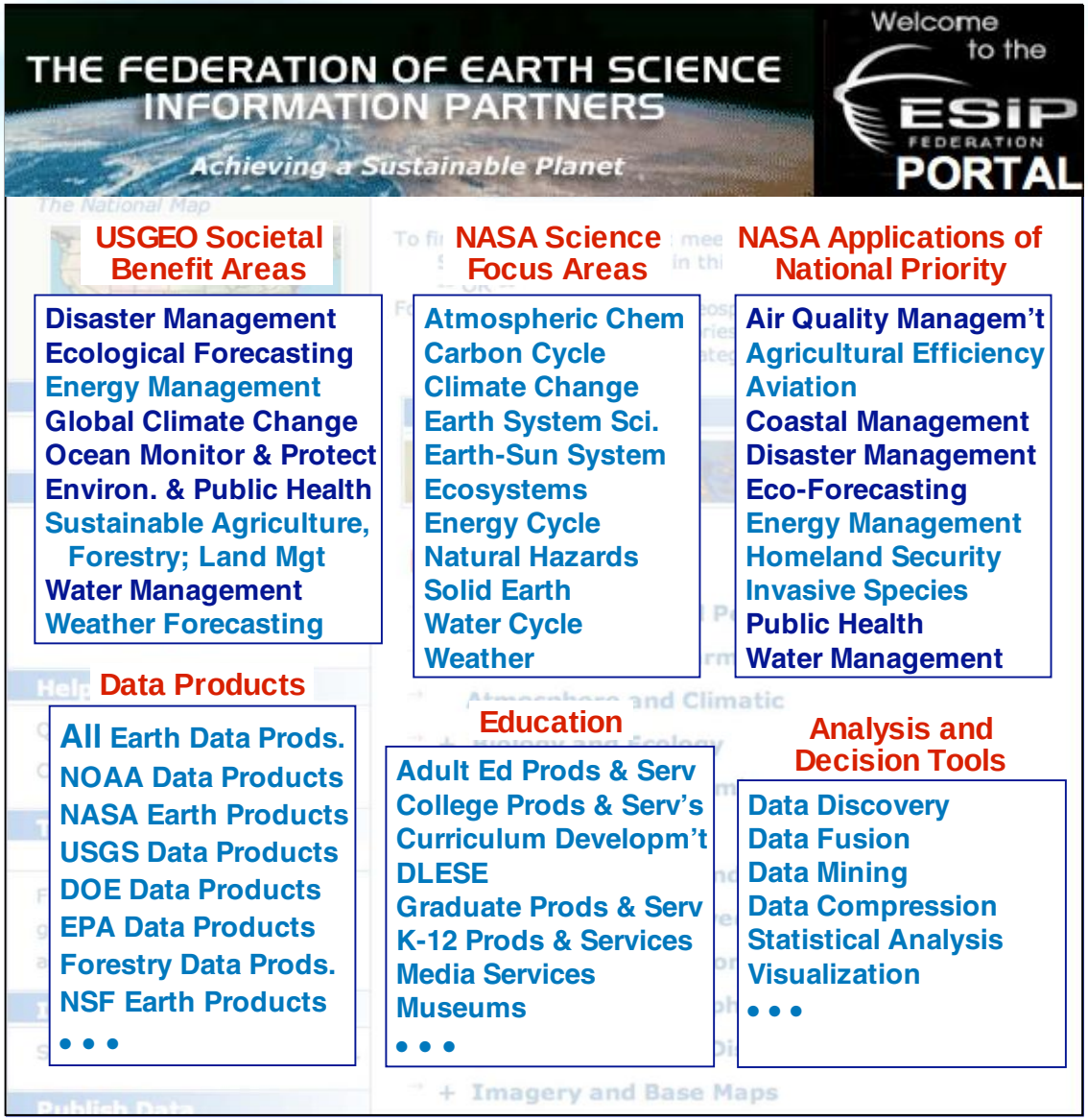
Usable Earth Science and Environmental Information

Rolling Out *GENESIS* in the Earth Information Exchange

The EIE will initially be hosted by Geospatial One-Stop

Rolling Out *GENESIS* in the Earth Information Exchange

The EIE will initially be hosted by Geospatial One-Stop



Yunck et al., 06/27/06

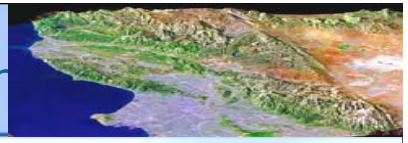
Entering the Global Computing Space

On-Demand Service Chaining

Portal Data Storage CPU Algorithms Networks

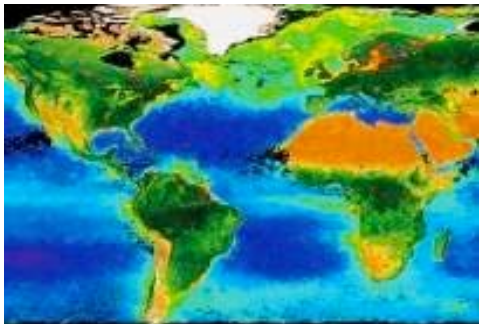
Come see a demonstration of Grid-based service chaining at the ESIP Federation Summer Meeting at Lamont-Doherty, Palisades, New York
July 19-21

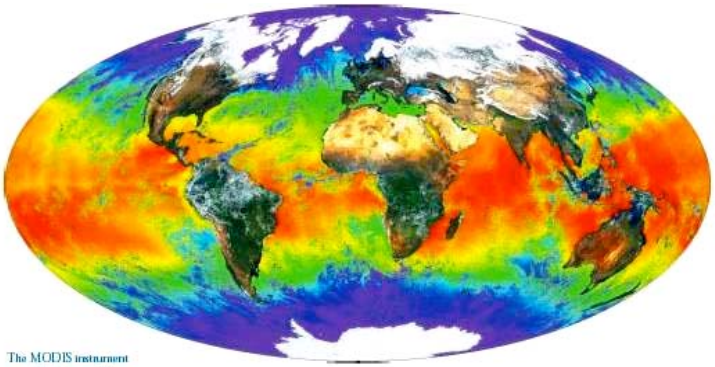
Or write to Tom Yunck at: tpy@jpl.nasa.gov



Summary: Directions and Themes

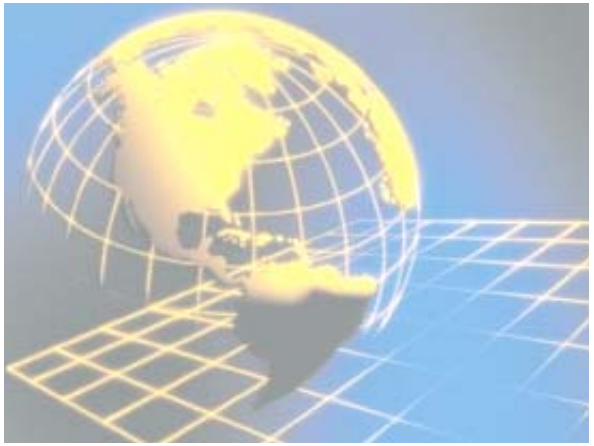
- Decentralized multiscale computing as a utility
- Peer-to-peer discovery, distribution, services
- Ubiquitous machine-machine communications
- Distributed control by XML messaging
- Semantic information representation
- Integrated modeling environments
- Visual workflow programming
- High-performance computing
- Standards Standards Standards





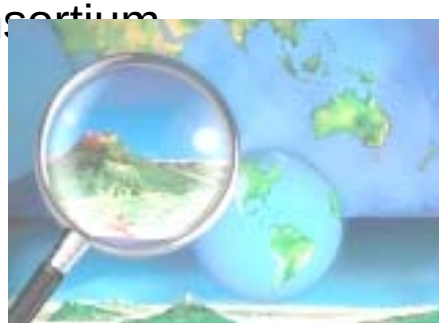
Guiding Themes

- Decentralization
  - The Global Grid
    - Peer-to-Peer
      - Machine-to-machine
        - Automated workflows
          - Distributed execution
            - Dynamic load balancing
              - Grid web services
                - Multi-scale integration
                  - Plug-and-play software



Data Discovery…

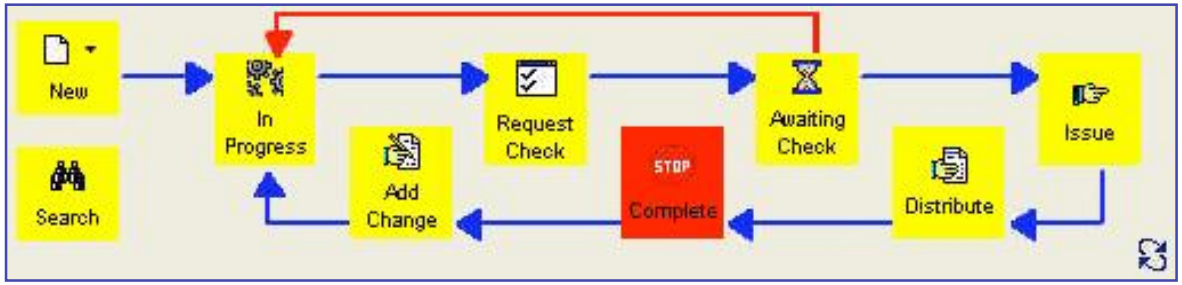
- Current:
  - ECHO, GCMD, FIND, EDG, SRB, UDDI, …
  - Integrated archives: EOS-Webster (UNH), GLCF (UMD), TRFIC (Mich State), GHRC (Huntsville), and many others
  - Co-op and commercial services: UNIDATA; Coop Climate Rainfall Data Center (CSU); Earth Data Discovery Consortium

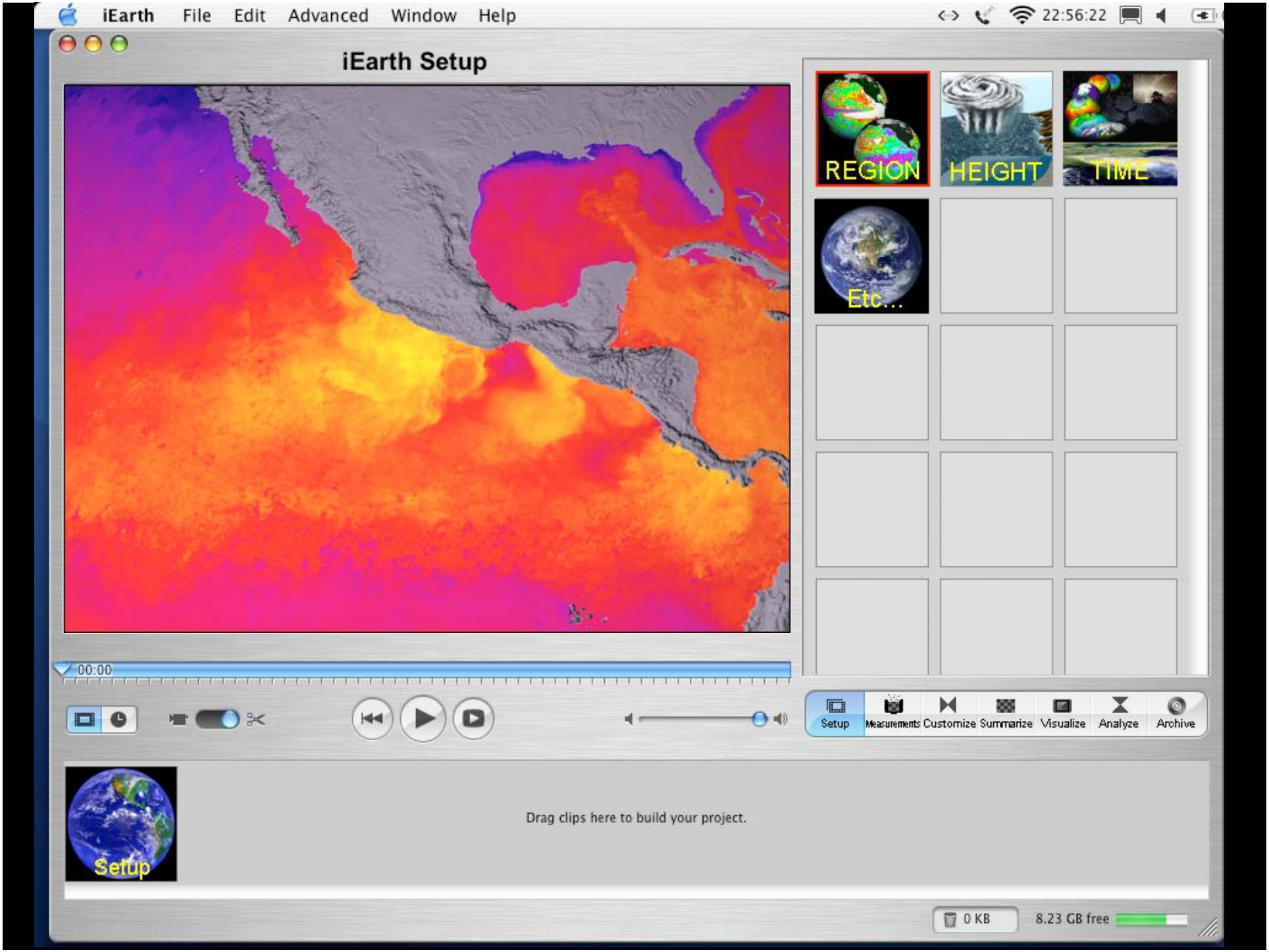


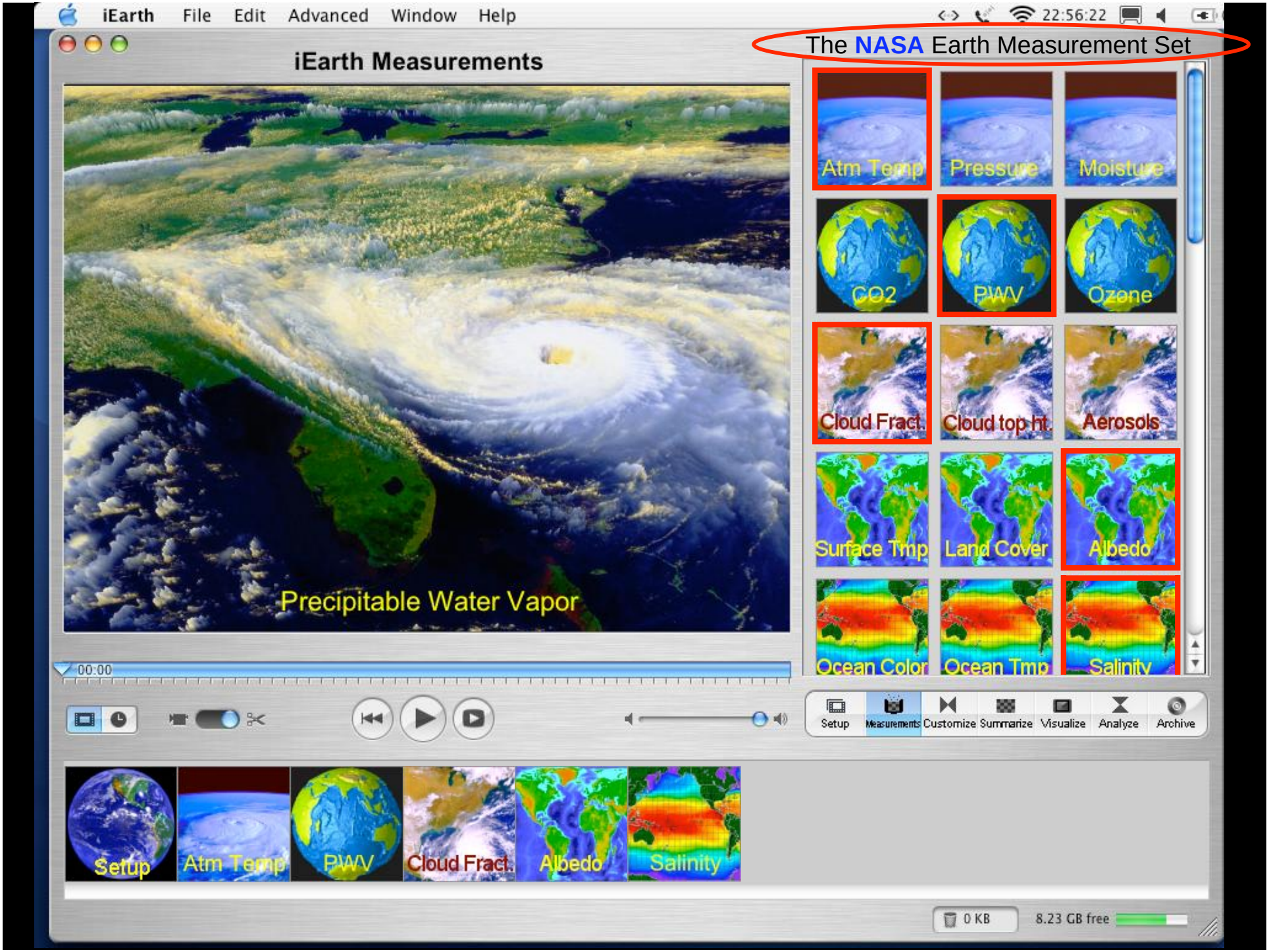
- Emerging:
  - Persistent archives and logical namespaces
  - Proliferating peer-to-peer exchange networks
  - Content-based-search / Semantic Web: Semantic representation of data enabling computers to understand web content.
  - Unidata THREDDS thematic data registries

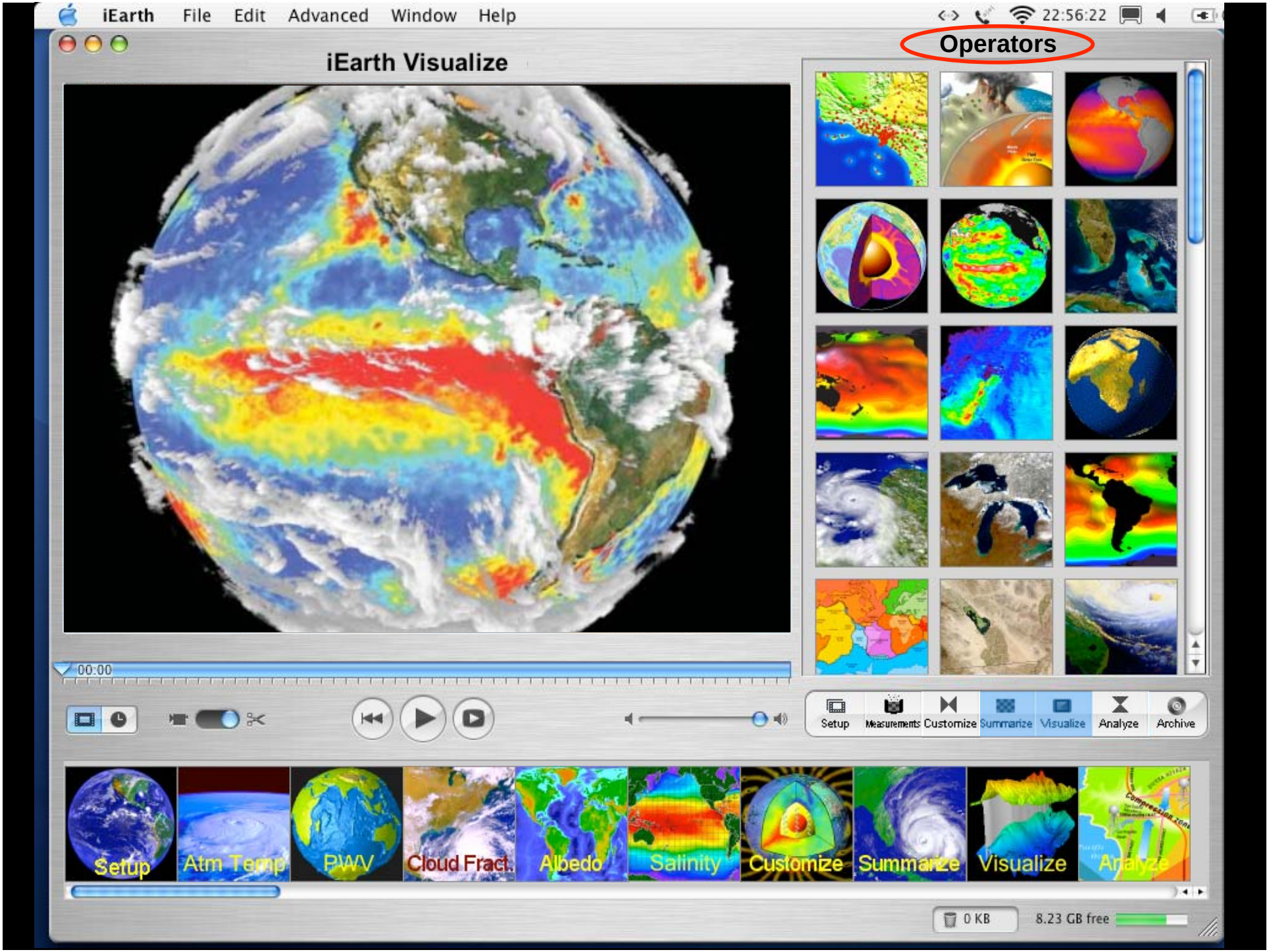
Automated Workflow Control

- "Invisible Hand" executive for loosely-coupled Grid: intelligent autonomous grid agents
- Dataflow documents, distributed execution engines
- Visual flow programming: drag'n'drop icons
- Common semantic framework – Semantic Web
- Automated parallelizing/workflow/resource allocation




iEarth File Edit Advanced Window Help
22:56:22
iEarth Setup
REGION
HEIGHT
TIME
Etc...
00:00
Setup
Measurements
Customize
Summarize
Visualize
Analyze
Archive
Setup
Drag clips here to build your project.
0 KB
8.23 GB free


iEarth File Edit Advanced Window Help
22:56:22
iEarth Measurements
The NASA Earth Measurement Set
Precipitable Water Vapor
Atm Temp
Pressure
Moisture
CO2
PWV
Ozone
Cloud Fract.
Cloud top ht.
Aerosols
Surface Tmp
Land Cover
Albedo
Ocean Color
Ocean Tmp
Salinity
00:00
Setup
Measurements
Customize
Summarize
Visualize
Analyze
Archive
Setup
Atm Temp
PWV
Cloud Fract.
Albedo
Salinity
0 KB
8.23 GB free


iEarth File Edit Advanced Window Help
22:56:22
iEarth Visualize
Operators
00:00
Setup
Measurements
Customize
Summarize
Visualize
Analyze
Archive
Setup
Atm Temp
PWV
Cloud Fract
Albedo
Salinity
Customize
Summarize
Visualize
Analyze
0 KB
8.23 GB free